Strong correlation between meiotic crossovers and haplotype structure in a 2.5-Mb region on the long arm of chromosome 21

Danielle M. Greenawalt,[1] Xiangfeng Cui,[1] Yujun Wu,[2] Yong Lin,[2] Hui-Yun Wang,[1] Minjie Luo,[1] Irina V. Tereshchenko,[1] Guohong Hu,[1] James Y. Li,[3] Yi Chu,[1] Marco A. Azaro,[1] Christina J. DeCoste,[4] Nyam-Osor Chimge,[1] Richeng Gao,[1] Li Shen,[1] Weichung J. Shih,[2] Kenneth Lange,[5] and Honghua Li[1,6]

[1]Department of Molecular Genetics, Microbiology and Immunology and [2]Department of Biometrics, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA; [3]Department of Computer Science, University of Maryland, Baltimore County, Baltimore, Maryland 21250, USA; [4]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA; [5]Department of Biomathematics, University of California Los Angeles School of Medicine, Los Angeles, California 90095, USA

Although the haplotype structure of the human genome has been studied in great detail, very little is known about the mechanisms underlying its formation. To investigate the role of meiotic recombination on haplotype block formation, single nucleotide polymorphisms were selected at a high density from a 2.5-Mb region of human chromosome 21. Direct analysis of meiotic recombination by high-throughput multiplex genotyping of 662 single sperm identifies 41 recombinants. The crossovers were nonrandomly distributed within 16 small areas. All, except one, of these crossovers fall in areas where the haplotype structure exhibits breakdown, displaying a strong statistically positive association between crossovers and haplotype block breaks. The data also indicate a particular clustered distribution of recombination hotspots within the region. This finding supports the hypothesis that meiotic recombination makes a primary contribution to haplotype block formation in the human genome.

[Supplemental material is available online at www.genome.org and http://www2.umdnj.edu/lilabweb/Publications. htm.]

It is known that in the human population, certain alleles of genetic markers within a short distance are in tight association (linkage disequilibrium or LD) and LD becomes weak or disappears when the markers are located farther apart (Ardlie et al. 2002). Chromosomal segments containing markers in LD are called haplotype blocks (Wall and Pritchard 2003). Haplotype blocks in the human genome were first described on a large scale for a 500-kb region of chromosome 5q31 (Daly et al. 2001) and the entire chromosome 21 (Patil et al. 2001), and subsequently in other regions of the genome (Gabriel et al. 2002; Twells et al. 2003; Olivier et al. 2004; Stenzel et al. 2004). Information on the haplotype structure of the human genome is of great interest because it can be used to significantly reduce the number of markers necessary for localizing genes responsible for complex diseases (Judson et al. 2002; Wang et al. 2002; Phillips et al. 2003). The progress of the Human HapMap Project has resulted in the comprehensive mapping of haplotype blocks across the entire human genome (The International HapMap Consortium 2003). However, very little is known about the mechanisms underlying the formation of haplotype blocks. It was noticed in the late 1990s that polymorphic frozen blocks (PFBs) were linked to form megabase haplotypes in the major histocompatibility com-

plex (MHC) region. Regions between these PFBs appear to contain localized recombination "hotspots" (Gaudieri et al. 1997), indicating that in the human genome recombination hotspots may have a primary role in LD breakdown.

There is a strong belief that meiotic recombination plays a primary role in shaping LD and therefore has a direct effect on the haplotype structure found in the human (Daly et al. 2001; Jeffreys et al. 2001, 2005; Cullen et al. 2002; Gabriel et al. 2002; Kauppi et al. 2003; Twells et al. 2003; Crawford et al. 2004; McVean et al. 2004). However, proving such a correlation requires direct evidence of the contribution of recombination on haplotype block formation. Computational approaches based on haplotype structure or LD in the human population have been shown to be helpful in determining recombination rates for either detailed analysis or on a genomic scale (Li and Stephens 2003; Crawford et al. 2004; McVean et al. 2004). However, recombination derived from the haplotype structure cannot, in turn, be used to study the correlation between recombination and haplotype block formation.

By using pooled sperm, a 216-kb segment in the class II region of the MHC was studied in detail to elucidate such a relation (Jeffreys et al. 2001). Six recombination hotspots were precisely located within regions where LD breaks down. However, since this study was performed in a relatively small and preselected region, the questions remain: Would the human genome contain recombination hotspots in a similar pattern and/or den-

sity; and do the hotspots fall between haplotype blocks, that is, within regions where LD breaks down in other regions? To address these issues another region, chromosome 1q42.3, which is more typical of the human genome, was analyzed (Jeffreys et al. 2005). The authors' approach allows them to focus on small subregions to learn a great deal about the mechanisms underlying meiotic recombination and its impact on the genetic structure of the human genome during evolution. However, by analysis of small regions, it is difficult to learn the distribution of recombination crossovers at levels higher than individual hotspots.

In-depth study of a large chromosomal region is also necessary for this purpose, but it is especially challenging because haplotype blocks are usually very small, ranging from several hundred bases to several hundred kb, and meaningful information cannot be obtained until a large number of meiotic products are scored with a high marker density. The diploid nature of the human genome and the difficulty in gathering pedigrees large enough to study make this analysis very challenging.

Single sperm typing (Li et al. 1988, for review, see Arnheim et al. 2003; Carrington and Cullen 2004) is an ideal approach for this study. Sperm cells are haploid and contain all genetic information about meiotic recombination. Practically, an unlimited number of sperm can be obtained from each donor. However, a sperm contains only a single copy of the genome and has been considered to be impractical for analysis with a large number of markers. This challenge has been overcome by our laboratory through development of a highly sensitive multiplex genotyping system that allows genotype determination of >1000 genetic markers in a single sperm (Wang et al. 2005).

Results

Identification of meiotic crossovers by single sperm typing

In the present study, a panel of 578 single nucleotide polymorphisms (SNPs) in a 2.5-Mb region on the long arm of chromosome 21 (from 38.01 Mb to 40.51 Mb) with an average intermarker distance of 4323 bp was selected from the public SNP database, dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html), maintained by the National Center for Biotechnology Information (NCBI). The markers were incorporated into our high-throughput genotyping system (Wang et al. 2005). Three Caucasian donors (D-8, D-11, and D-12) heterozygous at 131, 193, and 209 SNP loci, respectively, were selected for analysis. Single sperm from each donor were prepared by flow cytometry. After lysis, the polymorphic DNA sequences at all 578 SNP loci in each sperm were amplified with our high-throughput genotyping system. The resulting PCR product was used as a template for genotype determination on microarray by a single-base extension assay as described previously (Wang et al. 2005).

In total, 662 single sperm samples, 472 from D-11, 115 from D-12, and 75 from D-8 were genotyped at all 578 marker loci. Forty-one recombinants were identified from the 662 single sperm samples, each containing a single crossover. The crossovers identified represent a 6.19% recombination rate, 1.41 times the male average for chromosome 21, and 2.57 times the genomic average (Kong et al. 2002). As shown in Table 1, the size of these regions ranged from 2.6 kb to 98 kb, depending on the availability of informative markers, except one, X8, that was 306 kb.

Thirty recombinants (6.36%) out of 472 single sperm from donor D-11 were identified. The crossovers were found in 16 regions (X1 to X16) that are defined as being flanked by the nearest informative SNPs. Two additional individuals, D-12 and D-8, were genotyped to determine whether the pattern of crossover events occurring was similar between individuals. Six crossovers were identified among 115 sperm (5.22%) from individual D-12, and five among 75 sperm (6.67%) from individual D-8, as shown in Figure 1A and Table 1. These 11 crossovers fell into eight regions that were either located within or overlapped with (note, different donors have different sets of informative SNPs) regions from D-11 containing one or more crossovers.

When informative markers are used at a high density, each crossover will fall in an interval between the nearest informative markers, which is usually smaller than the expected interval size calculated based on the genomic or chromosomal averages, even if the crossovers occur randomly. Therefore, when a small region contains a single crossover, no matter how small this region is, it may not necessarily be a recombination hotspot. However, when more than one crossover is found in such a region, it may indicate a correlation between the occurrence of crossovers and the respective region. Six regions—X2, X3, X8, X10, X12, and X16—were found to contain more than one recombinant from D-11. The recombination rates in four of these regions are more than 10 times greater than expected based on the average rate of the 2.5-Mb region. However, such a comparison is very superficial because it does not take the probability of occurrence of these crossovers into consideration. For example, based on the regional average, the probability for one crossover to occur within the 48.5-kb region of X2 is 0.0012. Therefore, the probability of five or more crossovers occurring in this region should be 3.01×10^4 times the expected based on the binomial test. In total, six regions (X2, X3, X12, X16, X17, and X22) were found to have recombination rates significantly different than expected. Region X12, whose rate was 1.65×10^8 times the expected, was the highest (Table 1).

If the crossovers from all three individuals were considered additively, ten regions (X1, X2, X3, X4, X6, X8, X10, X12, X14, and X16) would contain greater than a single crossover. Six other regions contained only one crossover in each. However, these single-crossover regions do not weaken the nonrandomness of crossover distribution because, as shown by Jeffreys et al. (2001), the recombination rates between hotspots may vary in a wide range, up to a 260-fold difference. Therefore, regions with single crossovers can be considered as noninformative for proving crossovers occurring nonrandomly in these regions.

Correlation between regions containing crossovers and haplotype blocks

We examined whether the occurrence of crossovers is correlated with the haplotype block structure in the 2.5-Mb region. The chromosomal region flanked by X11 and X16 is only 298.16 kb, 11.9% of the entire 2.5-Mb region, but 12 (40.0%) crossovers fell into this region. The haplotype blocks, identified from the HapMap project (The International HapMap Consortium 2003), in this region are generally smaller (average size 5474 bp) than those in the entire 2.5-Mb region (average size 13,047 bp). In contrast, the regions between X2 and X3 and between X7 and X8 are 361.58 kb and 356.19 kb, respectively, 29% of the entire 2.5-Mb region, and are expected to contain 8.7 of the 30 crossovers but none were observed. The haplotype blocks in these two areas are generally larger, with an average size of 34.91 kb and 23.39 kb, respectively. The two largest haplotype blocks within

Table 1. Crossover areas identified in three donors from the 2.5-Mb region

Donor	Name	Start (bp)	End (bp)	Size (bp)	Number	Rate (cM)	Relative rate (cM/Mb^{-1})	Expected rate[a] (cM)	Observed/ expected	Probability of having observed crossovers or more[b]
		Region			**Crossovers**					
					Observed					
D-11	X1	38152612	38178652	26,040	1	0.2119	8.14	0.0645	3.28	—
	X2	38218211	38266729	48,518	5	1.0593	21.83	0.1202	8.81	3.0123E-04
	X3	38628305	38636185	7880	2	0.4237	53.77	0.0195	21.71	3.9856E-03
	X4	38739959	38758968	19,009	1	0.2119	11.15	0.0471	4.50	—
	X5	38797827	38811501	13,674	1	0.2119	15.49	0.0339	6.25	—
	X6	39032715	39093924	61,209	1	0.2119	3.46	0.1516	1.40	—
	X7	39262278	39272176	9898	1	0.2119	21.40	0.0245	8.64	—
	X8	39628371	39934305	305,934	3	0.6356	2.08	0.7579	0.84	6.9429E-01
	X9	39945953	39948593	2640	1	0.2119	80.25	0.0065	32.39	—
	X10	39991122	40023327	32,205	2	0.4237	13.16	0.0798	5.31	5.5319E-02
	X11	40168288	40189134	20,846	1	0.2119	10.16	0.0516	4.10	—
	X12	40196686	40202975	6289	5	1.0593	168.44	0.0156	67.99	1.6513E-08
	X13	40237524	40248168	10,644	1	0.2119	19.90	0.0264	8.03	—
	X14	40291433	40302607	11,174	1	0.2119	18.96	0.0277	7.65	—
	X15	40302607	40338244	35,637	1	0.2119	5.95	0.0883	2.40	—
	X16	40461817	40466449	4632	3	0.6356	137.22	0.0115	55.39	2.5274E-05
	Total			616,229	30	6.3559	10.31	1.5266	4.16	—
D-12	X17	38155214	38250530	95,316	2	1.7391	18.25	0.2361	7.37	3.0650E-02
	X18	38250530	38266729	16,199	1	0.8696	53.68	0.0401	21.67	—
	X19	39747011	39768010	20,999	1	0.8696	41.41	0.0520	16.72	—
	X20	40298992	40311185	12,193	1	0.8696	71.32	0.0302	28.79	—
	X21	40368736	40467182	98,446	1	0.8696	8.83	0.2439	3.57	—
	Total			243,153	6	5.2174	21.46	0.6024	8.66	—
D-8	X22	38202979	38250530	47,551	3	4.0000	84.12	0.1178	33.96	1.0359E-04
	X23	38739959	38761795	21,836	1	1.3333	61.06	0.0541	24.65	—
	X24	38985699	39039429	53,730	1	1.3333	24.82	0.1331	10.02	—
	Total			123,117	5	6.6667	54.15	0.3050	21.86	—
Total/Average		—	—	1,964,998	41	6.1934	2.48	—	—	—

[a]Values in this column are expected if the crossovers were evenly distributed in the 2.5-Mb region.
[b]Calculated by the binomial distribution test. *P*-values are the probability for the observed or more to occur and therefore to constitute a recombination hotspot. Because single crossovers cannot constitute any statistic power, no probability of occurrence can be calculated.

the 2.5-Mb region (176.53 kb and 194.74 kb) are also found in these two areas (Fig. 1A,B).

The correlation between the crossovers and haplotype block is also shown between the locations of the crossovers and haplotype block boundaries. Haplotype information for chromosome 21 is available on the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) based on the study by Patil et al. (2001) and can also be found at the Web site from the HapMap Project. We used a haplotype map constructed with genotypes from the HapMap Project for comparison (http://www.hapmap.org), because of the greater number of genotypes available, and because the map was constructed based on the genotypes of 90 CEPH individuals (Centre d'Etude de Polymorphism Humain, Utah Residents with Northern and Western European Ancestry) that are close to the genetic backgrounds of the donors used in the present study, while the map by Patil et al. (2001) used donors of mixed ethnicities. However, the structures of haplotypes in the crossover areas were similar for both maps. Although informative markers may not be located directly between haplotype block boundaries, as shown in Supplemental Figure 1, nine (X3, X4, X5, X9, X10, X11, X12, X15, and X16) of the 16 regions containing crossovers tend to occur between the haplotype block boundaries or within areas with less than half haplotype block coverage.

Between two informative markers there are three possible haplotype structures, which we defined as Types I, II, and III. A Type I region contained at least part of one haplotype block and the area between two haplotype blocks within which no haplotype structure is found. Type II regions were between two haplotype blocks and contained no haplotype structure. Type III regions were defined as areas within a single haplotype block (for graphical representation, see Supplemental Fig. 2). For in-depth analysis of the correlation between crossovers and haplotype blocks, the entire 2.5-Mb region was divided into subregions by the informative markers. Of the 2.5-Mb region, 1.87 Mb (74.7%) were covered by Type I regions, 0.11 Mb (4.4%) by Type II, and 0.52 Mb (20.9%) by Type III regions. Therefore, if the crossovers occurred randomly, the 30 crossovers from donor D-11 should be distributed as 22.41 in Type I, 1.32 in Type II, and 6.27 in Type III regions. In contrast, we observed 25 in Type I, 4 in Type II, and 1 in Type III, which is significantly different from the expected ($P = 0.005$ by Exact Test and $P = 0.006$ by Pearson's χ^2 Test.) These analyses revealed a strong association between the occurrence of crossovers and the areas between boundaries of haplotype blocks. However, because of the large number of Type I areas, which are less informative, more analysis was performed to further validate the correlation between haplotype blocks and crossovers.

Recombination hotspots may be present in clusters

The 2.5-Mb region displays a pattern in which most haplotype blocks <20 kb are clustered and the clusters are separated by one
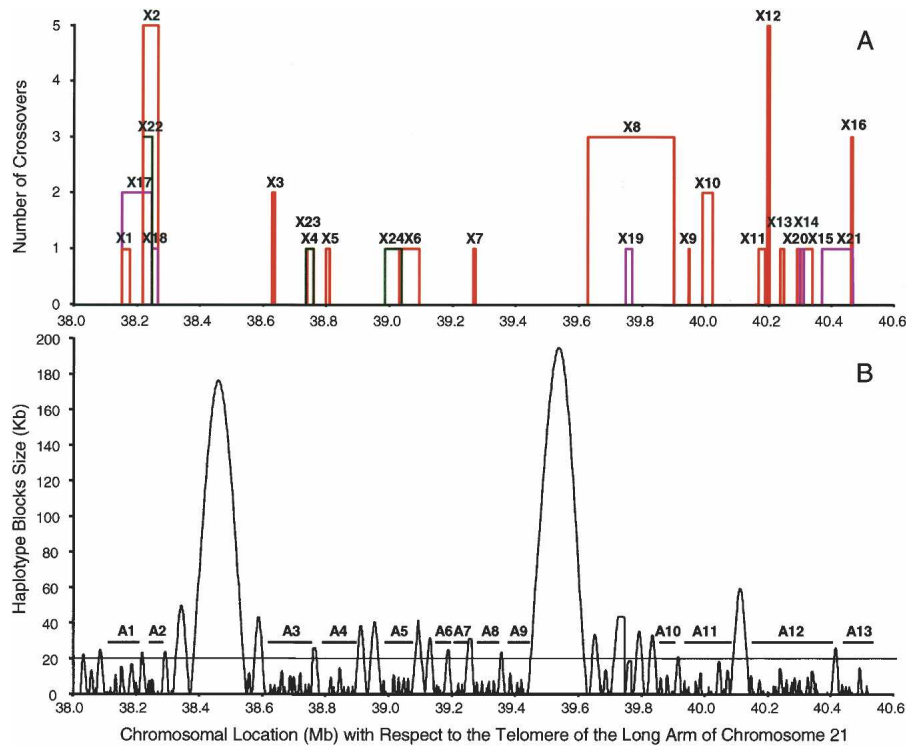
Figure 1. Schematic physical correlation between locations of crossovers and haplotype blocks in the 2.5-Mb region. (*A*) Distribution of recombination crossovers. Regions in which crossovers were identified are indicated as rectangles and labeled with an "X" and a number. The width of a rectangle shows the areas between informative markers, and height indicates the number of crossovers identified. Crossovers identified from individuals D-11, D-12, and D-8 are red, pink, and green, respectively. (*B*) Haplotype blocks and Areas A1–A13, containing only small blocks and spacers within the region. Haplotype blocks are peaks in dark blue with their bottoms indicating their locations and spanning areas. Areas with spacers in clusters are marked with horizontal bars and indicated with a letter "A" and a number.

(3) "spacers," which are the areas between the blocks, contain no haplotype block structure and occupy the remaining 0.79 Mb (31.6%) of the 2.5-Mb region. If the crossovers were randomly distributed, of the 41 crossovers observed, 17.55 should be found within the large blocks and 23.45 in either small blocks or within spacers. Instead, we observed 34.31 (83.68%) of the 41 crossovers within areas containing only small blocks and spacers. A Pearson's χ^2 Test produces a *P*-value of 0.0006, which reveals a strong association between the crossovers and the areas containing only small blocks and spacers.

Thirteen areas (A1–A13, Fig. 1B), between haplotype blocks >20 kb and containing small blocks (4 to 24 in each) and spacers, were analyzed in detail. As shown in Table 2, the average block size in these areas ranges from 3.0 to 7.8 kb. The average spacer size ranges from 3.5 to 8.8 kb, comparable with the block size in these areas. Other than these 13 areas, there are also smaller areas that may accommodate recombination, such as X19, which falls into a small area between two large blocks of 33.6 and 43 kb with a smaller block of 18.1 kb in between.

The number of crossovers in each area, A1–A13, was plotted against the number of spacers. As shown in Figure 2, the two parameters display a significant positive correlation ($R^2 = 0.64$), indicating a cumulative effect of the number of spacers on the number of crossovers. The correlation becomes very strong ($R^2 = 0.94$) if A2 is excluded.

Assuming that spacers, the region between haplotype blocks lacking any haplotype structure, contain recombination active elements (RAEs), the occurrence of recombination may be af-

or a few blocks >20 kb. To learn whether such a structure is correlated with the locations of the crossovers, the 2.5-Mb region was subdivided into three types of blocks based on their size: (1) large blocks, or haplotype blocks >20 kb, which occupy 1.07 Mb (42.8%) of the 2.5-Mb region; (2) small blocks, those blocks <20 kb, which account for 0.64 Mb (25.6%) of the 2.5-Mb region; and

Table 2. Features of blocks and spacers in the thirteen selected areas

| Area number | Area size | Block size (bp) | | | Spacer size (bp) | | | | Number of crossovers |
		Total	Average	Standard deviation	Total	Average	Standard deviation	Number	
1	109,147	55,840	6204	6531	53,307	5331	4128	10	2.61
2	47,819	16,841	5614	3555	30,978	7745	3841	4	5.80
3	145,041	76,665	6389	3852	68,376	5260	3590	13	3.71
4	111,689	40,928	5847	4693	70,761	8845	9931	8	1.00
5	96,113	46,907	7818	3314	49,206	7029	6245	7	1.65
6	29,666	12,037	3009	1827	17,269	3526	2820	5	0.00
7	38,747	12,690	4230	2771	26,047	6514	5058	4	0.00
8	76,668	32,885	5481	2716	43,783	6255	3502	7	0.21
9	70,786	29,450	7363	3632	41,336	8267	4522	5	0.00
10	52,113	21,619	5405	5196	30,494	6099	3454	5	0.51
11	156,106	73,501	7350	5550	82,605	7510	5731	11	3.63
12	260,046	117,684	5117	4127	142,362	5932	6996	24	10.41
13	89,340	36,530	4566	4483	52,810	6601	4576	8	3.40
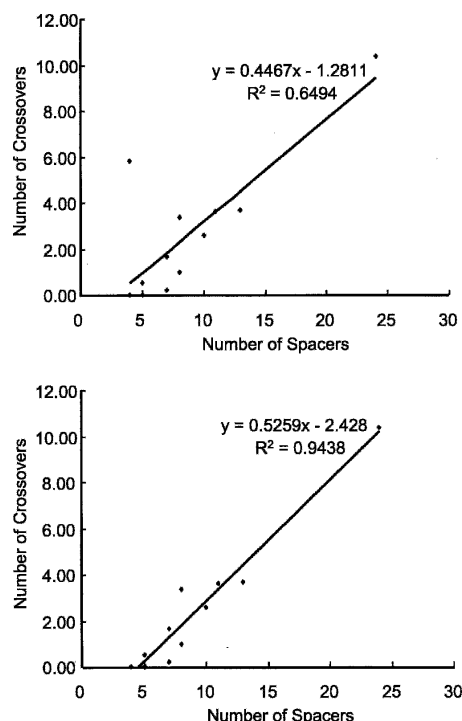
Figure 2. Correlation between the number of crossovers and the number of spacers in the 13 areas, A1–A13. (*Top*) Plot including A2; (*bottom*) plot excluding A2 from the correlation analysis.

fected by two factors: (1) the number of RAEs, and (2) the recombination activity of these RAEs. If the activities of the elements are similar, the number of crossovers should be directly proportional to the number of spacers, which is the situation in all areas except A2. Divergence of the number of crossovers in A2 from the linear function indicates that the elements in this area may be very active or A2 contains much denser RAEs than other regions. More studies would be necessary to learn whether such areas are common in the human genome.

Discussion

In a recombination study with pooled sperm, Jeffreys et al. (2001) identified six recombination hotspots within a 216-kb region. These spots are clustered in three regions, with 3, 2, and 1 hotspots in each. Hotspots in 'clusters' with greater than one hotspot are separated by haplotype blocks. Centers of hotspots in clusters with more than one hotspot are spaced 4010, 7970, and 3250 bp, which are within the distance range between the centers of two spacers described in the present study. A similar situation was found in their more recent study (Jeffreys et al. 2005). The spacers, regions between haplotype blocks containing no haplotype structure, described in the present study could be equivalent to the hotspots described in Jeffreys et al. (2005). However, 'clusters' of recombination hotspots analyzed in the present study are significantly larger and contain 4 to 24 hotspots (spacers) per cluster whereas only 1 to 3 per cluster were identified previously (Jeffreys et al. 2001, 2005). This could be explained by the fact that the recombination rate of the 2.5-Mb region analyzed is 2.57 times the genomic average, 2.97 times that (0.18 cM) of the 216-kb region studied (Jeffreys

et al. 2001), and 2.67 times that (0.20 cM) of the 200-kb region (Jeffreys et al. 2005). In addition, only 13 major clusters were found in the 2.5-Mb region (one cluster each 192 kb on average) whereas three were identified within 216 kb (one each 72 kb on average) (Jeffreys et al. 2001). Therefore, the size of the hotspot clusters described in the present study could be an explanation for the high recombination rate identified. In addition, the previous studies were focused on selected subregions and some recombination spots may have been missed.

The hypothesis that spacers analyzed in the present study are caused by recombination hotspots is consistent with current models of recombination (for review, see Keeney 2001) in which a double-strand break (DSB) occurs at the initiation sites of meiotic recombination. Migration of Holliday junctions away from initiation sites followed by junction resolution may result in crossovers at a distance from initiation sites. It has been shown in the yeast, mouse, and human (Keeney 2001; Kauppi et al. 2004) that crossovers are distributed in gradients flanking initiation sites. Because LD is formed during the long course of evolution, it would be difficult to learn the gradient of recombination based on observed haplotype structures. The size of spacers could be larger than the recombination hotspots detected experimentally. Over time, a few crossovers could shuffle the alleles of markers flanking initiation sites, breaking down the haplotype structure, even if the recombination rate is too low to detect experimentally at the ends of the crossover gradients.

In previous studies, the recombination rate at hotspots identified has varied (Jeffreys et al. 2001), whereas the number of crossovers identified in the present study is directly proportional to the number of spacers within clusters identified indicating similar activities of the hotspots. This may be explained by the fact that our analysis covers a large region of the chromosome, identifying a precise number of crossovers; however, fewer meiotic products were analyzed than in previous sperm typing studies (Jeffreys et al. 2001, 2005; Cullen et al. 2002). It is possible that only crossovers in regions with relatively high recombination activity were detected in our study. Therefore, the difference in recombination rates with respect to the number of spacers in each cluster is much smaller than that described previously (Jeffreys et al. 2001) and the cumulative effect of the number of spacers in each cluster is more obvious. The fact that the number of crossovers in A2 diverges from the linear function indicates that the recombination rates in some areas could be significantly higher than others. In addition, although extremes should be present among the clusters, considering that recombination rates in different hotspots should be different, when multiple spots are present in a single cluster, it is likely that the rate for the entire cluster evens out.

Our study provides direct evidence of the role of meiotic recombination on haplotype block formation. A much larger chromosomal region was studied in the present study than in previous analysis of recombination hotspots in sperm samples (Jeffreys et al. 2000, 2001, 2005; Cullen et al. 2002). Because haplotype maps of the human genome have been constructed (The International HapMap Consortium 2003, 2005), the information obtained from the present study may help researchers gain significant insight into the genetic structure of the entire human genome and the contribution of meiotic recombination to haplotype block formation.

Methods

SNP selection

A 2.5-Mb region from 38016911 to 40516122 bp of chromosome 21 was selected for analysis. Initially, 545 SNPs were selected with an intermarker distance of 4.65 kb from the NCBI dbSNP build 109. The locations of some selected markers changed during the span of the project with the update of human genome builds. Locations cited are based on NCBI Human Genome Build 34 version 3 and are consistent with mapping performed by sperm typing (http://www.ncbi.nlm.nih.gov/SNP/buildhistory.cgi). An additional 33 SNPs were added later from build 121 within X2, X8, and X12 of D-11. All SNPs in this region were analyzed to determine whether they would fit the constraints of our high-throughput multiplex PCR system (see Wang et al. 2005 for criteria). The RS numbers of the selected SNPs, and their corresponding primer and probe sequences are available online in Supplemental Table 1 (http://www2.umdnj.edu/lilabweb/Publications.htm).

Identification of informative individuals

Semen samples were the remains of infertility tests provided by Dr. David Seifer's endocrinology laboratory. All donors were shown to have normal fertility. Samples were sent to us anonymously with only the donors' ages and ethnicities labeled. There is no way for us to learn the donors' identities. Use of these samples was approved by the Internal Review Board at UMDNJ-Robert Wood Johnson Medical School. Samples were stored at −80°C. Twenty semen samples were genotyped in the same way as DNA samples described in Wang et al. (2005) following lysis of 0.5 µL of semen in 2.5 µL of lysis solution and neutralization (Pramanik and Li 2002) to identify donors with the largest informative fraction among the selected SNPs.

Single sperm preparation

Sperm from the selected semen samples were purified and stained for flow cytometric sorting as described previously (Pramanik and Li 2002). Sperm were sorted under sterile conditions into single wells of 96-well PCR plates with a Beckman Coulter ALTRA cell sorter, at the Rutgers University Flow Cytometry Facility, and a Becton Dickinson FacsVantage SE w/DiVa, at the Princeton University Flow Cytometry Facility, as described previously (Pramanik and Li 2002).

High-throughput multiplex genotyping

High-throughput multiplex genotyping was performed as described previously (Wang et al. 2005). Briefly, each sperm was lysed and neutralized followed by multiplex PCR amplification with all primers for the selected SNP panel. An aliquot of the multiplex PCR product was then used to amplify single-stranded DNA for hybridization, labeling, and detection on a microarray. Microarrays were scanned with a GenePix 4000B (Axon Instruments) microarray scanner at 10 micron per pixel, 100% power. The resulting images were digitized with GenePix Pro (Axon Instruments) to determine signal intensity. Genotypes were determined from signal intensities by a modified version of the genotyping program AccuTyping (G. Hu, in prep.) specific for sperm typing, developed in our laboratory. Haplotypes and recombination points were determined by comparing haplotypes from multiple sperm and semen samples.

Haplotype block analysis

A haplotype map constructed with genotypes from the HapMap Project was used for comparison. Genotypes for the CEPH trios in the HapMap database (http://www.hapmap.org) were saved to text files and were used for haplotype construction with the Haploview program (Barrett et al. 2005).

Statistical cluster analysis

Of the 41 crossovers, 22 were precisely located within the areas containing only small blocks and spacers. The remaining 19 were found in regions containing large blocks, small blocks, and spacers. Crossovers in these regions were allocated based on the assumption that crossovers are distributed evenly in the 2.5-Mb region. For example, X2 had 5 crossovers. A total of 32.1% of this region was in an area with a large block and 67.6% in a neighboring area with small blocks and spacers. Therefore, we considered 1.6 crossovers were located in the area with the large block and 3.4 crossovers were in the area with small blocks and spacers. The result would be biased toward allocating more crossovers to the areas with large blocks than it should be since all available data so far indicate that crossovers are correlated with spacer regions.

Acknowledgments

References

Ardlie, K.G., Kruglyak, L., and Seielstad, M. 2002. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **3:** 299–309.

Arnheim, N., Calabrese, P., and Nordborg, M. 2003. Hot and cold spots of recombination in the human genome: The reason we should find them and how this can be achieved. *Am. J. Hum. Genet.* **73:** 5–16.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21:** 263–265.

Carrington, M. and Cullen, M. 2004. Justified chauvinism: Advances in defining meiotic recombination through sperm typing. *Trends Genet.* **20:** 196–205.

Crawford, D.C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M.J., Nickerson, D.A., and Stephens, M. 2004. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* **36:** 700–706.

Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G., and Carrington, M. 2002. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.* **71:** 759–776.

Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J., and Lander, E.S. 2001. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29:** 229–232.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. *Science* **296:** 2225–2229.

Gaudieri, S., Leelayuwat, C., Tay, G.K., Townend, D.C., and Dawkins, R.L. 1997. The major histocompatability complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *J. Mol. Evol.* **45:** 17–23.

The International HapMap Consortium. 2003. The International HapMap Project. *Nature* **426:** 789–796.

———. 2005. A haplotype map of the human genome. *Nature* **437:** 1299–1320.

Jeffreys, A.J., Ritchie, A., and Neumann, R. 2000. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Mol. Genet.* **9:** 725–733.

Jeffreys, A.J., Kauppi, L., and Neumann, R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* **29:** 217–222.

Jeffreys, A.J., Neumann, R., Panayi, M., Myers, S., and Donnelly, P. 2005. Human recombination hot spots hidden in regions of strong marker association. *Nat. Genet.* **37:** 601–606.

Judson, R., Salisbury, B., Schneider, J., Windemuth, W., and Stephens, J.C. 2002. How many SNPs does a genome-wide haplotype map require? *Pharmacogenomics* **3:** 379–391.

Kauppi, L., Sajantila, A., and Jeffreys, A.J. 2003. Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.* **12:** 33–40.

Kauppi, L., Jeffreys, A.J., and Keeney, S. 2004. Where the crossovers are: Recombination distributions in mammals. *Nat. Rev. Genet.* **5:** 413–424.

Keeney, S. 2001. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52:** 1–53.

Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* **31:** 241–247.

Li, N. and Stephens, M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165:** 2213–2233.

Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* **335:** 414–417.

McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* **304:** 581–584.

Olivier, M., Wang, X., Cole, R., Gau, B., Kim, J., Rubin, E.M., and Pennacchio, L.A. 2004. Haplotype analysis of the apolipoprotein gene cluster on human chromosome 11. *Genomics* **83:** 912–923.

Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., et al. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294:** 1719–1723.

Phillips, M.S., Lawrence, R., Sachidanandam, R., Morris, A.P., Balding, D.J., Donaldson, M.A., Studebaker, J.F., Ankener, W.M., Alfisi, S.V., Kuo, F.S., et al. 2003. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat. Genet.* **33:** 382–387.

Pramanik, S. and Li, H. 2002. Direct detection of insertion/deletion polymorphisms in an autosomal region by analyzing high-density markers in individual spermatozoa. *Am. J. Hum. Genet.* **71:** 1342–1352.

Stenzel, A., Lu, T., Koch, W.A., Hampe, J., Guenther, S.M., De La Vega, F.M., Krawczak, M., and Schreiber, S. 2004. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum. Genet.* **114:** 377–385.

Twells, R.C., Mein, C.A., Phillips, M.S., Hess, J.F., Veijola, R., Gilbey, M., Bright, M., Metzker, M., Lie, B.A., Kingsnorth, A., et al. 2003. Haplotype structure, LD blocks, and uneven recombination within the LRP5 gene. *Genome Res.* **13:** 845–855.

Wall, J.D. and Pritchard, J.K. 2003. Haplotype blocks and linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **4:** 587–597.

Wang, N., Akey, J.M., Zhang, K., Chakraborty, R., and Jin, L. 2002. Distribution of recombination crossovers and the origin of haplotype blocks: The interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.* **71:** 1227–1234.

Wang, H.Y., Luo, M., Tereshchenko, I.V., Frikker, D.M., Cui, X., Li, J.Y., Hu, G., Chu, Y., Azaro, M.A., Lin, Y., et al. 2005. A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. *Genome Res.* **15:** 276–283.